access

NCSA | University of Illinois | Vol 23 No 2 | Summer 2010

NCSA

access

Vol 23 No 2 Summer 2010

On the Web

News from NCSA is always available on the Web. Sign up for our RSS feed or email service for the latest developments from NCSA and its partners.
www.ncsa.illinois.edu

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
1205 West Clark Street
Urbana, IL 61801
217-244-0072

Who we are

The University of Illinois at Urbana-Champaign's National Center for Supercomputing Applications (NCSA), one of the five original centers in the National Science Foundation's Supercomputer Centers Program, opened its doors in January 1986. Over the years NCSA has contributed significantly to the birth and growth of the worldwide cyberinfrastructure for science and engineering, operating some of the world's most powerful supercomputers and developing the software infrastructure needed to efficiently use them.

That tradition continues as the center, Illinois, IBM, and their partners in the Great Lakes Consortium for Petascale Computation develop what is expected to be the first computer dedicated to open scientific research capable of sustaining more than one petaflop, or one quadrillion calculations per second. Called Blue Waters, the system will come online in 2011. It will be dedicated to massive simulations and data analysis projects that will improve our society, health, environment, and economic competitiveness. NCSA and the consortium will also work with research communities to create the new software technologies, scientific applications, and educational programs needed to take full advantage of this new system.

Blue Waters will benefit from NCSA's ongoing focus on cyberenvironments, cyber-resources, and innovative systems research. Cyberenvironments give research communities the means to fully exploit the extraordinary resources available on the internet (computing systems, data sources and stores, and tools). Cyber-resources ensure computing, data, and networking resources are available to solve the most demanding science and engineering problems and that the solutions are obtained in a timely manner. Innovative systems research involves testing and evaluating the performance of emerging computing systems for scientific and engineering applications.

In addition, NCSA is a key partner in the National Science Foundation's TeraGrid project, a $150 million effort to offer researchers remote access to some of the fastest unclassified supercomputers as well as an unparalleled array of visualization tools, application software, sensors and instruments, and mass storage devices.

The center also leaves its mark through the development of networking, visualization, storage, data management, data mining, and collaboration software. The prime example of this influence is NCSA Mosaic, which was the first graphical Web browser widely available to the general public. NCSA visualizations, meanwhile, have been a part of productions by the likes of PBS's NOVA and the Discovery Channel. Through its Private Sector Program, top researchers explore the newest hardware and software, virtual prototyping, visualization, networking, and data mining to help U.S. industries maintain a competitive edge in the global economy.

Support for NCSA is provided by the National Science Foundation, the state of Illinois, industrial partners, and other federal agencies. For more information, see www.ncsa.illinois.edu.

On the cover

Less than two years after breaking ground, the University of Illinois' National Petascale Computing Facility (NPCF) was completed and ready for occupancy in June. The facility will be home for the Blue Waters sustained-petascale supercomputer when it comes online next year. The building will also house other computing, networking, and data systems. NPCF is an 88,000-square-foot facility of which 20,000 square feet is machine room. It will achieve LEED Gold certification, a benchmark for the design, construction, and operation of green buildings. NPCF's forecasted power usage effectiveness (PUE) rating is an impressive 1.1 to 1.2, while a typical data center rating is 1.4. PUE is determined by dividing the amount of power entering a data center by the power used to run the computer infrastructure within it, so efficiency is greater as the quotient decreases toward 1.

Summer Issue Design:
CarltonBruettDesign

access contents

An Expert Opinion

Building bridges from research to education

FEW OF THE ADVANTAGES of the national cyberinfrastructure will be realized without scientists and engineers who understand the new capabilities that it provides.

The cybereducation team at NCSA helps bridge the gap between research and education by conducting activities and programs to disseminate advanced technologies to new—as well as established—communities, encouraging them to adopt computational technologies as learning tools. The group conducts a wide variety of programs, from middle school through college graduate level.

One of our initiatives is the Institute for Chemistry Literacy through Computational Science (ICLCS). A five-year, $5 million Math Science Partnership program, ICLCS targets 118 rural school districts and 124 high school chemistry teachers in Illinois. Teachers in the program spend two weeks for each of three summers at NCSA, learning how to use computational resources and tools to transform their students' science education.

In much the same way that researchers are using computational and collaboration tools to advance science and share best practices, ICLCS has shown that the use of computational tools in the classroom combined with support for a virtual professional learning community (VPLC) leads to a statistically significant increase in achievement for high school students in rural schools based upon standardized testing. In addition, findings to date indicate that with each year of participation in the program teachers are continuing to gain chemistry content knowledge at the highest levels. Since over 30 percent of these teachers are the only science teacher in their small district, many feel that for the first time in their career they have science colleagues with whom to communicate and share ideas. With over 24,480 logins, 6,916 postings, and over 340,000 views since summer 2007 among teachers, faculty and researchers, the VLPC is well on its way to being a self-sustaining community of practice.

With support from an NSF Robert Noyce Master Teaching supplement, we are providing a cohort of ICLCS teachers with a unique infrastructure of support as they go through the National Board Certification process.

Our ICLCS successes are being noticed. The ICLCS team was invited to present at the 2009 U.S. Senate Technology Showcase and to a selected group of the U.S. House of Representatives in January 2010. In addition, ICLCS is changing the way chemistry is taught at the University of Illinois at Urbana-Champaign, where we are partnering with chemistry instructors and faculty to develop computational labs to support existing wet labs for general chemistry classes, which serve 3,500-4,000 each semester, and to deploy an online organic chemistry course using visualization tools and WebMO to construct 3D molecules.

WebMO is a web-based computational quantum chemistry software package used by ICLCS teachers, Illinois chemistry professors, and their students. Over the course of 15 months, nearly 32,000 jobs had been submitted by teachers (19,000) and by their students (13,000).

Recently, the NCSA cybereducation team deployed the first National Science Olympiad event using computer modeling. The Science Olympiad National Tournament was held on the Illinois campus this past May for nearly 1,500 students from 49 states. Ninety-two middle school and 50 high school students representing the top competitors in their state competed in the ModelThis! competition, using Vensim and NetLogo to answer questions related to predator-prey interactions and epidemiology, respectively.

One of NCSA cybereducation's longest-running programs is Girls Engaged in Mathematics and Science (GEMS). It was created in 1994 to encourage middle school girls to consider mathematics- and science-oriented careers, to gain confidence in doing math and science, and to take advanced-level mathematics and science courses in high school. Since its inception, the GEMS program has provided over 1,200 middle school girls with access to leading-edge research of female role models and to computational tools. The first GEMS website was created by a team of eighth-grade girls using NCSA Mosaic. Girls even participate in grid activities to connect with researchers and other girls in countries around the world.

The work of the cybereducation team would not be possible without our education interns, many of whom spend multiple years at NCSA and have gone on to create Farmville and MergeFM as young entrepreneurs. Our interns come from many colleges and departments on campus. We make a concerted effort to recruit interns, who are paid from grant funds, and have engaged many students from underrepresented communities over the years. In addition, eight interns are participating in a research experience for undergraduates program with an NCSA researcher.

Interns conduct a wide variety of activities, including developing instructional materials for education communities that infuse technology and authentic research tools into pre-college and undergraduate courses, and evaluation activities that assess the impact of our activities.

Interns also collaborated with Millersville University to develop a set of new features to existing LEAD-2-Learn modules for the Linked Environments for Atmospheric Discovery project. Interns created an interactive, inquiry-based, and self-guided experience for undergraduate education. They also created a modified version of Unidata's Integrated Data Viewer (IDV), called Basic IDV, for novice users that reduced the number of parameters needed to conduct activities. The result was greater accessibility using the minimally configured computers often found in secondary schools. Unidata has since adapted this feature in IDV.

The best part of leading the cybereducation team is working with creative and innovative students at all levels. They are having a big impact on science education.

Edee Wiziecki
Cybereducation Lead
NCSA

A better understanding

Building the Blue Waters sustained-petascale supercomputer also requires developing the applications that will allow researchers to do their science on the machine. Bill Gropp, the chief applications architect for Blue Waters, tells Access' Barbara Jewett what that work involves and how it feeds into the future of high-performance computing.

Q: Let's start with Blue Waters; tell me about your role in that.

A: I have several roles in Blue Waters. I am one of the co-PIs so I'm involved in monitoring the project and participating in various ways in what we're doing and what we need to do, with particular focus on the software. I'm also supervising a couple of projects that are looking at pieces that will contribute to that. One project I'm leading is porting libraries that either some of the PRAC-approved applications use or, in a couple of cases, libraries that we expect applications to need. PRAC is the National Science Foundation's Petascale Computing Resource Allocation program.

Q: How long does it take to develop a library?

A: Well, it depends on how much you need to do. If what you want to do is just port and see what trouble you have, that doesn't really take very long. If you really want to tune the library to make good use of the hardware, then that can take months.

We also are looking at I/O libraries. At large scale, approaches that work on a thousand nodes, such as having every processor write out its own file, is not a good way to use the I/O system when you have 10,000 nodes and hundreds of thousands of cores. It's also not a good way to do your analysis, with hundreds of thousands of files—some applications have each process write out one file per variable, and now we're talking about hundreds of billions of files; there are better ways. There's no real reason to do it that way other than in the past most I/O systems were not that capable of doing better, and if you only had a few hundred or a few thousand files you could get away with doing your I/O that way. There are I/O libraries that are up to the task of providing the application developers with workable alternatives. That's one component of what we try to do—help the applications developers.

The other project that we're looking at is support for programming models. We like to provide people with alternatives so they can make good use of the system. The most obvious is the hybrid programming model that mixes MPI with threads—either OpenMP or pthreads. And we'll also be providing good quality parallel languages that support what's called the partitioned global address space (PGAS) programming model. One of these is an extension to C called Unified Parallel C or UPC. So that you can define, for example, an array that is distributed across the parallel machine and easily access any part of that array. If this distributed data structure fits your application, it actually makes it easier to write different types of codes. Also, because it's a language, the compiler can perform optimizations that a library can't. IBM is doing this with their xlupc compiler and they already have some nice results. Of course, we don't expect users to rewrite their applications in a new language, so it will be possible to mix programming models in a single application. For example, you can write a module in UPC and link it with your MPI program. We'd like to see applications consider some performance-critical piece of code that would be suitable for UPC and try it out, linking with their MPI-based application.

Q: Is that something new and novel? It sounds like something not commonly done.

A: It's not commonly done. UPC has been around for quite a while, there are some users, there's a book on it, there are a number of groups that have built compilers and so forth, and it's been used on other supercomputers, but it hasn't really become mainstream. And I'm not sure it will become mainstream, but we want to make it easier and more feasible for groups to experiment with it and see if it will help solve some of their problems.

I wouldn't say that this is the first time support for UPC is available, but I would say this is the first time any supercomputing center has made it a goal that UPC would be available and interoperable with other parallel programming models, including MPI. We can use it to help better understand how we can move past the current MPI barrier in a way that meets the needs of applications.

I'd be the first one to say that it is relatively easy to define a new programming model that is exactly what you need for your application. The reason that MPI has been so successful is that the MPI programming model has been what almost everyone has needed to do their science. It is harder than people sometimes think to create a programming model that is applicable to everyone. By making it possible for people to experiment with other approaches without having to start over, we're probably going to give them programming alternatives that emphasize parallelism. We're going to give everyone, not just developers, a programming model better suited to where we go next.

Questions & Answers

Q: As we go forward with exascale, the codes are going to have to change, aren't they?

A: Yes, an exascale system is going to have far more concurrency than the petascale systems. We're looking at 300,000 cores at least for a sustained petascale system. At exascale you're more likely to be looking at tens of millions of threads of concurrency, and you're going to have to manage those in some way. And you are reaching scalability problems. If you think about the fact that messages between different parts of the machine will take different amounts of time to arrive because nodes are not the same distance apart, you start ending up with variations in timing that at a thousand nodes are really irrelevant but at a million nodes can become very important. And in the typical MPI model, programs assume that they have control of all cores all the time, and that each phase of the computation and communication will take the same amount of time.

But with exascale you start having a more dynamic response to when things get scheduled, when things get done. You can do that in MPI but it is not as easy, and there is no reason to still have to do that programming. So we need to start looking for models that will help us forge that track without sacrificing some of the things that MPI has given us in the ability to in fact express—the algorithms, providing information on performance-critical issues like data locality and so forth. So that we can gain more experience on how we put these things together.

I think we will see more hybrid models of programming so that we don't use one model everywhere. Many programmers don't like the hybrid programming model because it is hard to use. But the reason that the hybrid model is hard to use is that the parts have not been designed to fit well together. So the real problem is not that it is a hybrid model, the problem is that it is not a well-designed hybrid model.

An exascale system is likely to be hybrid hardware, partly because an exascale system will have to be maybe two orders of magnitude more power efficient than Blue Waters—and Blue Waters is pretty power efficient. To do that, we are not going to be able afford general-purpose cores. We might be able to have lower-powered general-purpose cores, but then we have to have even greater concurrency because that's your tradeoff.

So what I think we will see is that the exascale system is likely to have heterogeneous hardware because it will be hardware that is specialized for say, control flow, hardware that is specialized for streams of data, hardware that is specialized for vector operations that are different than what you get with streams; there might be hardware that is specialized for minimizing data motion. And you'll have to program all these things. And all that sounds pretty frightening, but to do it having a uniform programming model that hides everything from the user won't work. So what will work, then, is a programming model that tries to minimize the programmer's pain and makes it as easy as possible for the programmer to work with the different hardware components. I think that is not as bad as some people might think. And we have some beginning experience with this with the work that is going on here and elsewhere on the use of GPUs. I don't expect an exascale system to have GPUs attached to nodes like the current systems, but I wouldn't be surprised if the features of GPUs don't become part of what is inside an exascale processor chip. It won't be an extra chip on the side, it will be within the processor chip. The software and the tools that we're starting to develop will help us understand what we need to know to use that part of an exascale system.

Q: What are some other things that we're doing here at NCSA and the University of Illinois that go along with this?

A: There's a lot that is going on. There's fundamental work in computer architecture for the hardware that's going on between the computer science and electrical and computer engineering departments. There's work with programming models, there's work with tools to help you understand the performance that you're getting. One of the other problems to date has been that to a large extent a lot of the software work has been an art rather than engineering. In many applications, there is a trial and error approach to improving performance. So one of the things we are doing as part of the Blue Waters project and in these departments is trying to develop a better understanding of performance, better tools for modeling the performance of applications, better tools for applying the transformations needed to improve performance. With any aspect of engineering there is always an art to it, but it needs to be more systematic, more quantitative.

And so with the Blue Waters project we have several groups helping to develop analytic models of their applications' performance, which can help guide us and identify where the biggest gap is between the performance you should be getting and the performance you are getting. There are other efforts that are looking at tools that could be applied across the whole application so if, for example, you want to change a data structure, you can change it everywhere in your application. Such changes are something that computers are good at. And

Wen-mei Hwu, another co-PI and GPU expert, is developing tools for using GPUs so you can use analytic models and understand how a code should perform. It also helps identify where there are greater bottlenecks and that affects things and how you may design the code. And their tools will become ever more important as the architecture becomes more complex and more specialized.

Q: Some people say that by the time the Blue Waters operations end in 2016, we'll be ready for exascale, others that we'll be working with petascale for a very long time. Do you want to gaze into your crystal ball and say when we might be moving on?

A: When operation of Blue Waters reaches its predicted end, we won't be at exascale. Exascale is roughly 100 times faster than Blue Waters, and I just don't think we'll be there. Certainly not with the kind of architecture that we're looking at now.

I think that exascale is possible. There's a lot to do in terms of hardware research and software research. But I think we will go through another turn of an intermediate system, probably around 2015–2016, that might be a 100 petaflop. There are people who would like to see an exascale system by 2018. That's very aggressive. It is doable, but you'll have to sacrifice either cost or power.

Sometimes I think Blue Waters may be the last homogeneous general purpose processor system, because to get much past this without more power and without a lot more money you are going to have to give up something. You might give up the homogeneous parts and that would allow you to put a lot more computational power into the same footprint and the same power envelope, but it would be essentially like having a new edition, a new system. But that is going to require the right software tools. At exascale there is a risk of spending all of your energy moving data, not actually doing computations with it. And that will require building algorithms that are very limited in the data they would move, much more so than we do now.

Q: How does IACAT play into some of the things we're doing here at Illinois?

A: The Institute for Advanced Computing Applications and Technologies provides a way to connect NCSA staff with the rest of campus, looking at some of these questions. We have a couple of projects that are looking in general at advanced computing—not all of them are looking at exascale—but three of them that are looking at the petascale to exascale issues. One of those is looking at use of computers in applications. So that brings the knowledge here, and the access and the expertise in the applications, to the work that's being done in GPU systems, so those tools can actually be applied to real applications in petascale situations. There's another project that is doing similar things, but within a different programming model, that provides a more dynamic approach to the use of several of these processors to address the problems of systems that evolve over time. And again, working with real applications as compared to working with benchmarking and testing codes. And there's a third project that is looking more at algorithms with a focus on multiscale problems. To get to exascale or to get to trans-petascale then they'll need to re-think the algorithms. We need different algorithms, not for the code problems, but for the problems we'll put on an exascale system. And those problems tend to have many components and parts.

Q: We're taking about extreme-scale machines, but not everybody has an extreme-scale type of problem to solve. For those researchers who need computational power to solve their problems, but are not Blue Waters or extreme-scale users, what sort of computing resources will be available in the future for them?

A: Well, even though they won't be using exascale they'll benefit from solving the power problem for it. The processing cores are not going to get much faster, so the only way you make a processor faster is if you provide more parallelism on it. Your laptop, in a couple of years, might have 32, 64, or even 128 processor cores on it. Doubling will be the only way to get more computational power. And even now I think all laptops have at least two cores, so everybody has to deal with parallel processing. A lot of the techniques that we are developing make use of different levels of parallelism. And when you are looking at parallelism on Blue Waters, it's not equal threads; there is a hierarchy to the parallelism. So we have to understand how to make use of the eight cores on each chip and the 32 cores on each module. All that work will, in a few years time, be of good use in your laptop.

Q: So we'll all benefit from Blue Waters.

A: We'll all benefit from a better understanding of how to make good use of specific levels of parallelism. Just the very top level of parallelism atop the whole thing is something that only people who have the most demanding problems will have to worry about. But the tools that we're developing will help the whole software stack. Sometimes when looking at the big picture it gives you a better way to understand how to solve individual pieces. □

By J. William Bell

Getting the rabbit in the hat

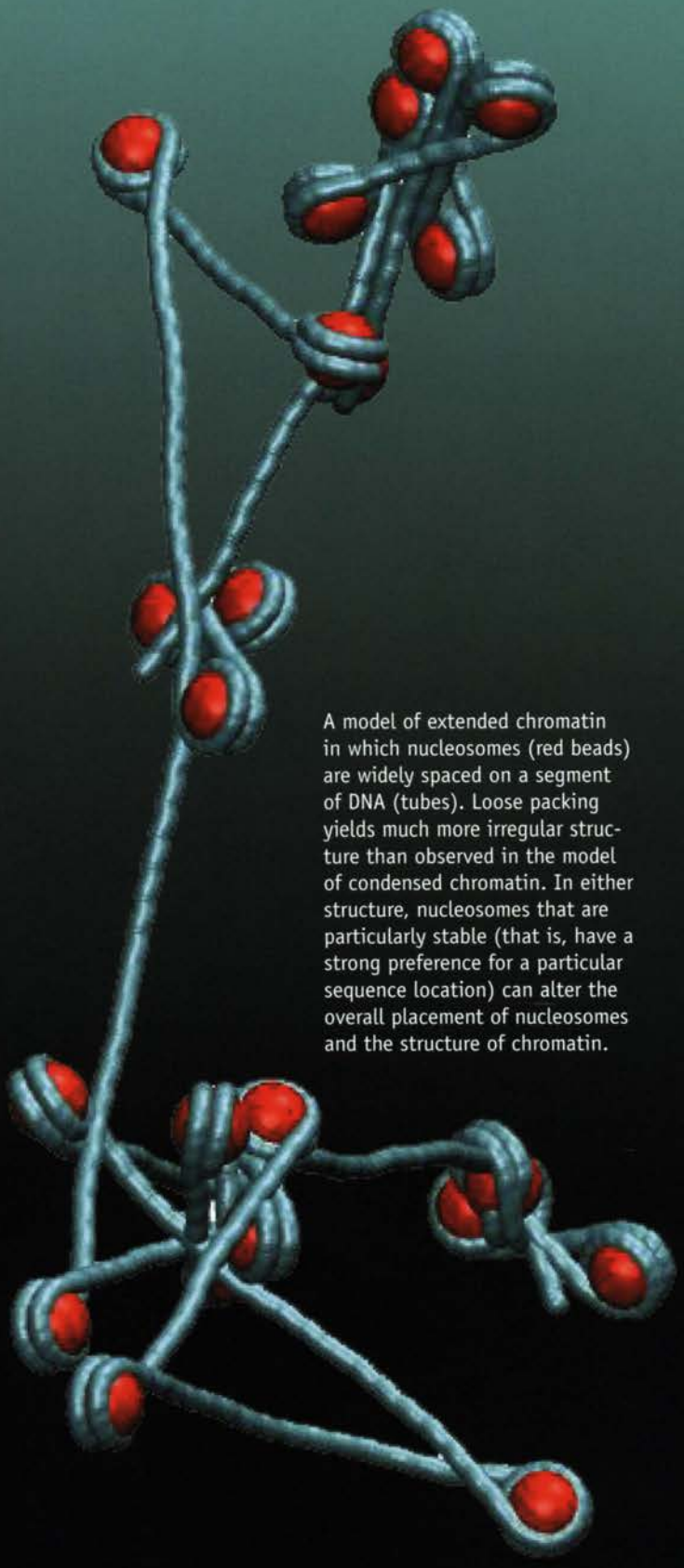
Tulane University researchers investigate how tiny units of genetic material fold into long fibers—and impact some of our bodies' most fundamental processes.

A model of condensed chromatin in which the nucleosomes (blue beads) are closely packed on a length of DNA (tubes). Tight packing limits the range of possible locations for any single nucleosome and thus the fiber assumes a comparatively regular structure.

THE HUMAN GENOME'S three billion base pairs, if stretched into a straight chain, would be about one meter long. Yet, by folding itself up, it fits in only a fraction of the volume of a cell's nucleus. Some very clever origami is required to keep the genome functional in this state instead of just becoming a tangled mess.

But that folding operation—in which short segments of DNA are folded into nucleosomes forming a fiber-like structure called chromatin—is more than just a means of sneaking a rabbit into a very, very small hat. All genomic processes are somehow impacted by which segments of DNA are wrapped into nucleosomes, the order in which nucleosomes line up, and which face of the nucleosomes' DNA is directed in and which is directed out.

A team led by Tulane University's Tom Bishop is beginning to use NCSA's Abe supercomputer, as well as supercomputers throughout the TeraGrid, to investigate some of the most basic features of how nucleosomes behave.



A model of extended chromatin in which nucleosomes (red beads) are widely spaced on a segment of DNA (tubes). Loose packing yields much more irregular structure than observed in the model of condensed chromatin. In either structure, nucleosomes that are particularly stable (that is, have a strong preference for a particular sequence location) can alter the overall placement of nucleosomes and the structure of chromatin.

Most people are familiar with the double helix structure of DNA. They're also familiar with the structure at the other end of the spectrum, when DNA is tightly packaged into chromosomes for segregation to the daughter cells during division. This structure can be seen with a simple microscope. "Somewhere between these two extremes is how DNA exists during most of the cell cycle, thus our interests in nucleosomes and chromatin," according to Bishop, a research associate professor at Tulane's Center for Computational Science.

"At this level of folding is where the cellular machinery spends most of its time doing its work based on the DNA blueprints." He says to think of this work as the "the three Rs and a T of molecular biology"—replication, regulation, repair, and transcription. "Understand these mechanisms and you've covered the fundamentals of molecular biology. There's still much to learn about these mechanisms, but nucleosome stability is likely relevant to all. Our current interest is how stability relates to promoting or restricting the expression of proteins."

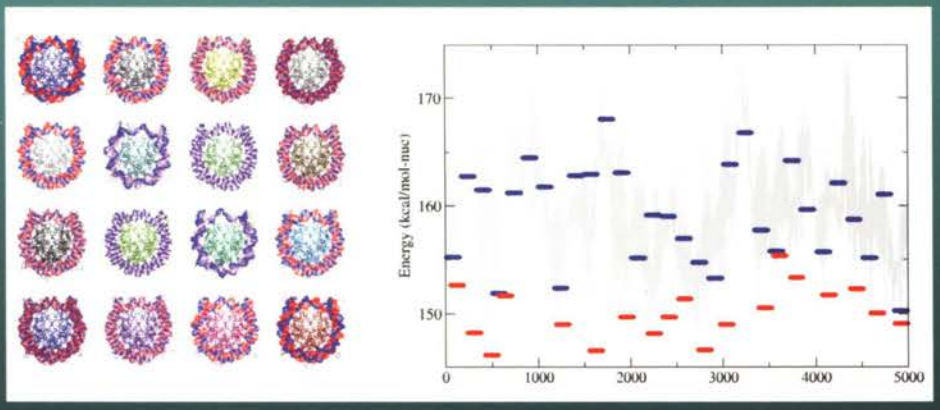
With energy

When DNA folds into fibers of chromatin, the underlying nucleosome structures are all remarkably similar. One hundred and forty-seven base pairs form a spiral staircase around eight proteins called histones. Genetics requires that the sequence vary. Packing requires that the histones be somewhat impartial to the sequence being packed Nonetheless, sequence differences impact nucleosome behavior.

"They may look identical, but different amounts of energy are required to make each nucleosome," says Bishop.

This variation in energy determines, in part, the placement of nucleosomes in chromatin fibers and how likely the nucleosomes are to move about in the fiber. It also determines how much energy is required for a cell's machinery to gain access to DNA inside a given nucleosome.

Nucleosomes with low energy thresholds can more easily be displaced, yielding direct, open access to the DNA blueprints. Nucleosomes with a high energy threshold allow more limited access. Stable nucleosomes perform other tasks. For example, they can help organize the local structure of chromatin and are often associated with promoter complexes that direct the cell machinery toward transcription start sites for particular genes.



Nucleosomes, left, are the fundamental structural unit of chromatin. Each of the 16 nucleosomes pictured has a different sequence of DNA and therefore a different free energy of formation. For a continuous segment of DNA, this results in a continuous energy landscape as a function of sequence (grey line, middle) in which nucleosome positions appear as discrete entities. If only the lowest energy conformations (red footprints, middle) are occupied, the resulting chromatin is extended. To achieve condensed chromatin (top,right) requires that some nucleosomes occupy high energy locations (blue footprints, middle).

Bishop's team is currently investigating the energy profile of thousands of nucleosomes in the yeast genome using both all-atom and coarse-grain computational models. They're determining how variations in the DNA sequence inside a nucleosome influence its energy profile and thus its position and stability in the chromatin fiber.

"A lot of effort has been devoted to protein folding, an amino acid origami problem," Bishop explains. "We're trying to understand chromatin folding, a nucleic acid origami problem." □

Project at a glance

Current team members
Tom Bishop
Rajib Mukherjee

More information
http://dna.ccs.tulane.edu/

Funding
National Institutes of Health
National Science Foundation

This effort also greatly benefits from the work of Tevfik Kosar, the Petashare Resource, and Shantenu Sha, all at Louisiana State University. The efforts of the NAMD developers are critical as well. Louisiana Optical Network Initiative (LONI) resources were used to demonstrate proof of concept.

Access online
www.ncsa.illinois.edu/News/Stories/nucleosomes

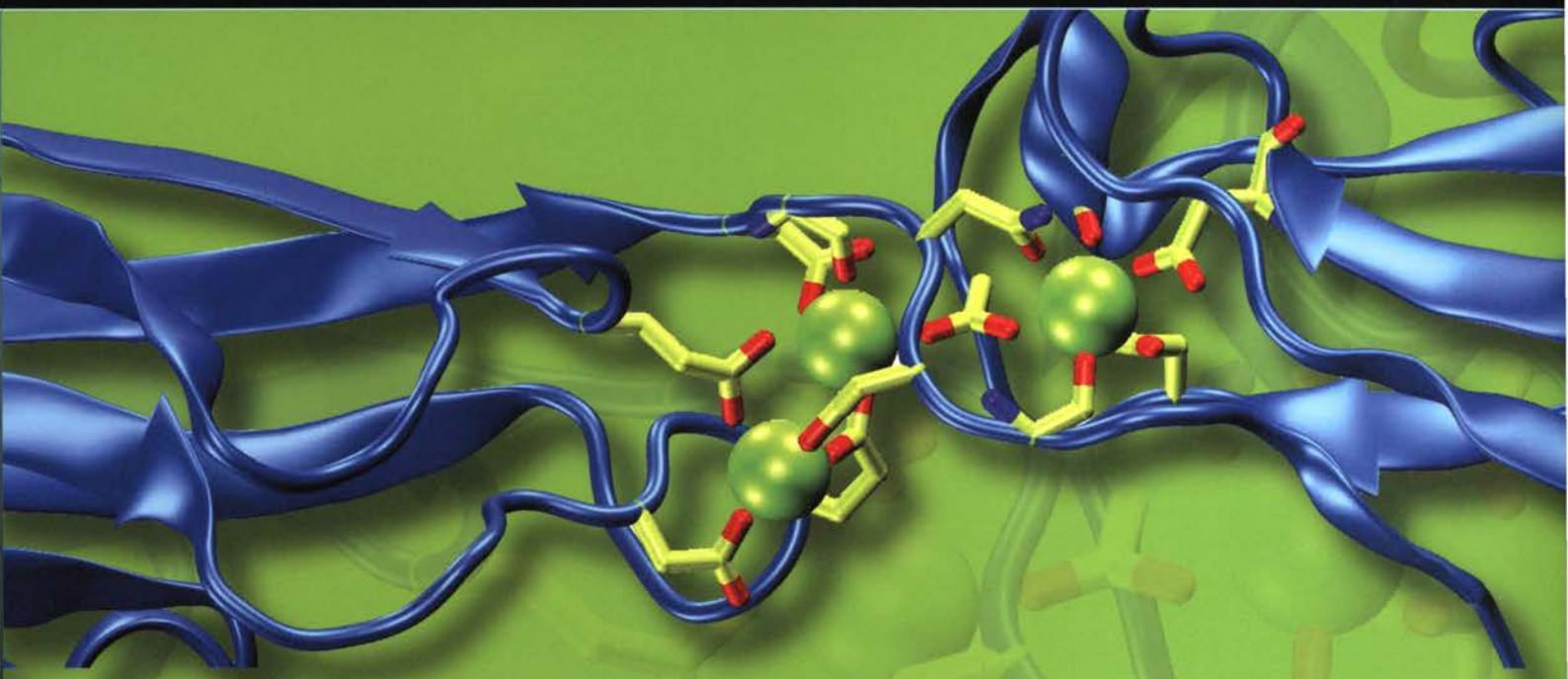
A Blue Waters problem

Tom Bishop's team models nucleosomes with a simulation code called NAMD. Bishop was one of the code's earliest users as a graduate student in Klaus Schulten's Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign.

As part of the National Science Foundation's Petascale Computing Resource Allocations program, NAMD is currently being scaled to run on NCSA's Blue Waters sustained-petascale supercomputer.

Blue Waters will come online in 2011. According to Bishop, more powerful machines like Blue Waters will allow scientists like him to go from simulating single nucleosomes to simulating ensembles of mononucleosomes or short segements of chromatin. Those sorts of simulations will provide further insights into the nucleosome's behavior and it's impact on replication, regulation, repair, and transcription.

"The smallest piece of chromatin is ginormous from an atomic viewpoint; its heterogeneous and sensitive to its solvent environment. To capture such heterogeneities and the solvent environment requires millions of atoms. That's an ideal problem for Blue Waters, " Bishop says.



TeraGrid resources help Harvard team gain insights into deafness

Adapted from a Harvard Medical School article

PROTEINS ARE THE MACHINERY behind cellular processes in living organisms, and much has been learned about how they function in the perception of some external stimuli from their three-dimensional structures. However, little is known about the protein molecules involved in touch and auditory function. Now, a team from Harvard University has resolved the atomic-level structure of a protein essential for sound perception. The structure was then simulated using TeraGrid supercomputers to determine how it functions in hearing and deafness. The research was published in the April 15 issue of *Neuron*.

Sound becomes an electrical signal in hair cells of the inner ear. These cells are sensitive mechanoreceptors that are moved by sound, stretching fine filaments called "tip links" to pull open ion channels and generate electrical signals that are sent to the brain. The tip link is made of two non-classical cadherin molecules, cadherin-23 and protocadherin-15, which are defective in some forms of hereditary deafness.

A team led by Rachelle Gaudet and David P. Corey used X-ray crystallography to determine the molecular structure of cadherin-23's tip, with and without a mutation causing deafness. Yet, static images of the protein provided little information about its function in hearing mechanics, so the team decided to test cadherin-23's elasticity using molecular dynamics simulations on TeraGrid systems at NCSA and the Texas Advanced Computing Center (TACC).

The protein structures were modeled in water boxes to mimic their biological environment; the simulation systems encompassed up to 355,000 atoms. Using the NAMD software developed by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign, the team performed hundreds of molecular dynamics simulations using the Ranger system at TACC and the Abe cluster at NCSA. The powerful machines and large number of processors available allowed them to test the elasticity of cadherin-23's tip in multiple near-physiological conditions. Analysis of more than 5 terabytes of data generated by the simulations revealed an essential role for calcium ions in the mechanics of cadherin-23.

The protein was found to be stiff in the presence of calcium, but weak when calcium was not bound to it. A deafness-causing mutation was found to alter calcium binding to the cadherin-23 protein, thus reducing its mechanical strength and possibly making tip links prone to rupture, thereby impairing sound perception. The simulations performed at TeraGrid facilities allowed the team to get insights about the role of calcium ions and cadherin-23 mechanics in hearing and deafness that cannot be obtained with any other technique, and provided predictions that can be experimentally tested.

Project at a glance

Team members

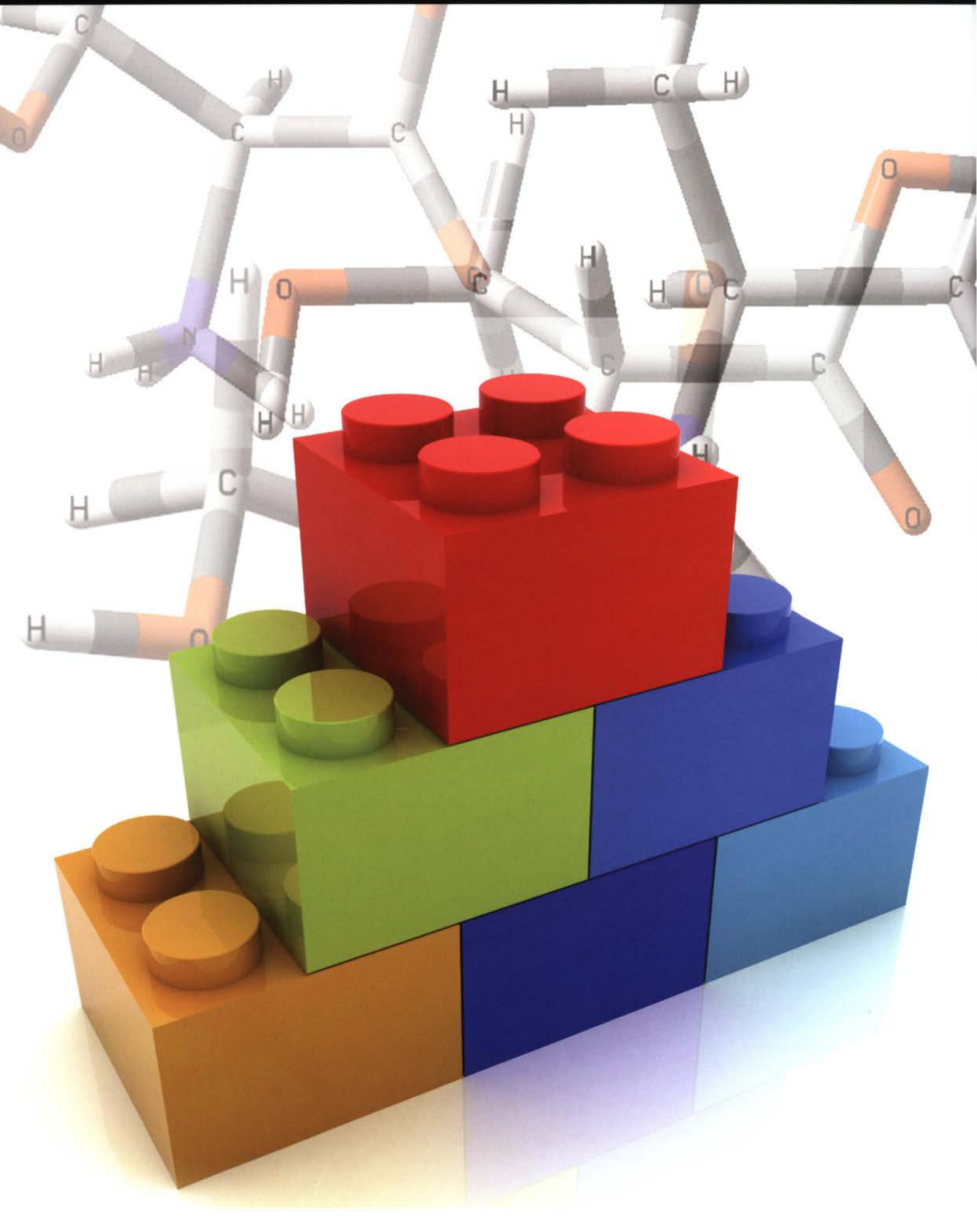
David P. Corey
Rachelle Gaudet
Marcos Sotomayor
Wilhelm A. Weihofen

Access online

www.ncsa.illinois.edu/News/Stories/deaf

Left to right: Rachelle Gaudet, David Corey, Marcos Sotomayor, Wilhelm Weihofen. (Photo by Josh Touster)

"The outstanding resources and support provided by the TeraGrid were essential to our work," says Marcos Sotomayor, a member of the research team at Harvard Medical School. "We were particularly happy with Ranger and Abe, as we were able to use NAMD on both machines extensively and smoothly. We also benefited from NCSA's Unitree storage system, which was essential for us as we were able to recover all our data when one of our local disks failed." □



by Barbara Jewett

Inspiring minds

"Having access to NCSA kept my projects alive!"

ALTHOUGH VARIATIONS OF the gratefulness theme are frequently expressed when researchers share information about their work conducted using NCSA resources, Mihaela Bojin's appreciation goes a little deeper than most. Without access to free supercomputing clusters, she most likely would have abandoned her amino acid research for several years.

Like many young researchers, a few years ago Bojin faced an important decision: how to balance a research career's demands with those of a young family. Her post-doctoral research position was coming to a close and she had a toddler. She wasn't sure the rigors of academic research to build a career at a major university would afford her the amount of time she wanted to spend with her child, even with the assistance of a supportive and helpful husband. She opted instead for an assistant professor position in the chemistry department at Queensborough Community College (QCC) in Bayside, New York, part of the City University of New York system.

Queensborough in general—and the chemistry department in particular—has a research focus not typically found at community colleges. The chemistry department even boasts a new laboratory devoted solely to faculty/student research projects. But as is the case with so many colleges and universities, QCC lacks research compute power. For a computational scientist like Bojin, this is an obstacle. Enter NCSA.

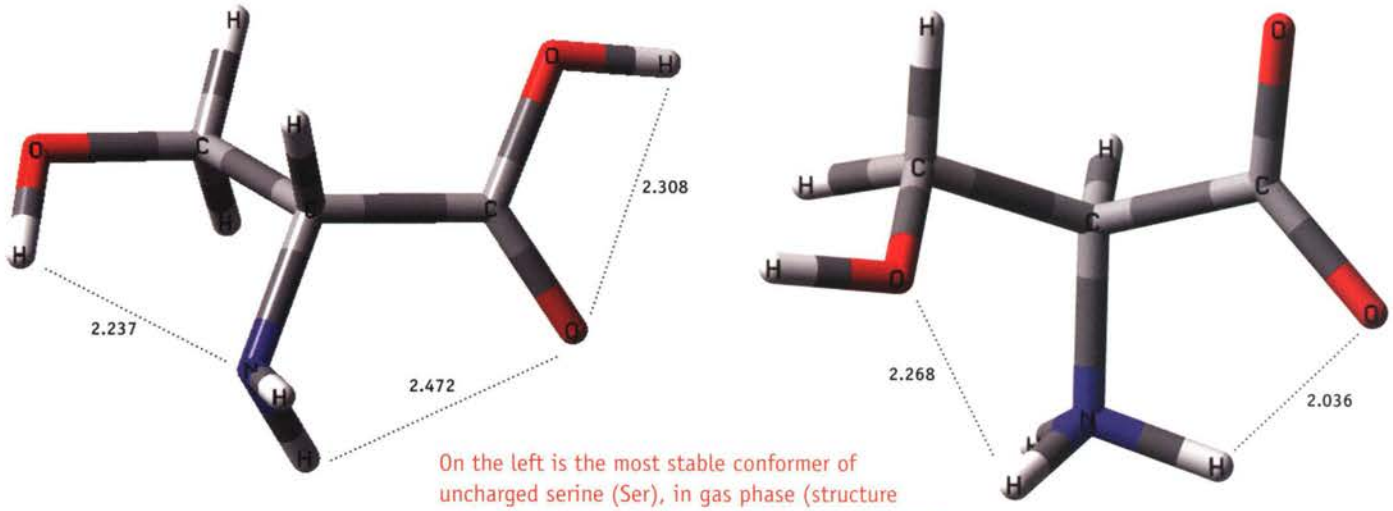
Exposing undergraduates

Bojin joined Queensborough in 2007, knowing she wanted to expose her students to computational science. In fact, she started training students her very first semester as a professor, even though she had no funds.

"I had them read articles and we would discuss what is computational chemistry in very general terms. And luckily it crossed my mind," says Bojin, "that I used to have a grant from NCSA when I was a post-doc a few years before I joined here, and I remembered that it was fairly easy to get a grant. I thought, 'I need to apply for that to get my project started.' So I applied for a grant and I got it very fast. And then we actually got to run calculations and try our ideas, to try them out on the clusters. So for me, the program has been super essential. I don't know how I can be more grateful for it. It has literally been my start-up funding."

Bojin and her undergraduate team of two students are using NCSA's Abe and Cobalt clusters. Her teams have also run on the now-retired Tungsten. They are exploring hydrogen bonding in amino acids, specifically what happens to those bonds when the acidity or basicity of the environment changes. Think of a protein as being a Lego, she says. Amino acids are just the little building blocks in the Lego.

"I am looking at each of these small pieces to see how many forms they can take, because each of them is actually very flexible," she explains. "So they can move and they can form distinct conformations."



On the left is the most stable conformer of uncharged serine (Ser), in gas phase (structure retained in water), and on the right its zwitterionic form, in water (not located in gas phase).

When you change the acidity of the medium of the amino acid, how does its internal bonding change? How influenced are they by the medium? These are questions Bojin hopes to answer.

The fundamental question

"Amino acids are part of proteins, and proteins are part of our body. So it is very important to know how proteins work to figure out how healthy we are," says Bojin. "We are thinking that if we go to the smallest part of the protein, these amino acids, then maybe we can expand our knowledge to the bigger proteins. How proteins are changed when the acidity of the medium changes. It has been proved that slight changes in the medium in the protein can lead to mutations, and mutations can lead to, for example, cancers and all kinds of diseases. So it is important to know about all of these changes, then you can learn maybe how to prevent them or how to treat something like this.

"That is very far from our actual research. Our research is more on the fundamental level. The ultimate long-term goal is to correlate these changes in acidity for the small amino acids with what happens in an actual protein."

Amino acids' intra- and intermolecular interactions are significantly influenced by surrounding residues and the pH of their environments. Hydrogen bonds, in particular, control folding in secondary and tertiary structures in proteins and significantly affect enzymatic activity, Bojin explains. By employing density functional methods her team calculated different conformations of asparagine, arginine, glutamic acid, serine, and threonine in their neutral (no charges), zwitterionic, acidic (protonated), and basic (deprotonated) forms, in gaseous and aqueous media. They found that changes in acidity critically influenced and limited hydrogen bonding patterns, and thus the stability of the resulting conformers. The team is currently working on glutamine, aspartic acid, histidine, and dipeptidic conformers.

Bojin presented her research at the fall national meeting of the American Chemical Society (ACS) in both 2008 and 2009. Four undergraduate students also gave presentations at each meeting, and two more will participate the 240th ACS meeting in Boston in August. Bojin's students also presented their work at the ACS Mid-Atlantic Regional Meeting in 2008 and 2010, and at the Collegiate Science & Technology Entry Program (CSTEP) meeting in 2009 and 2010. At the latter her student Daniel Sangobanwo won first prize in the poster competition. In addition, her students give presentations every year at honors conferences and undergraduate research symposia of the ACS-New York section.

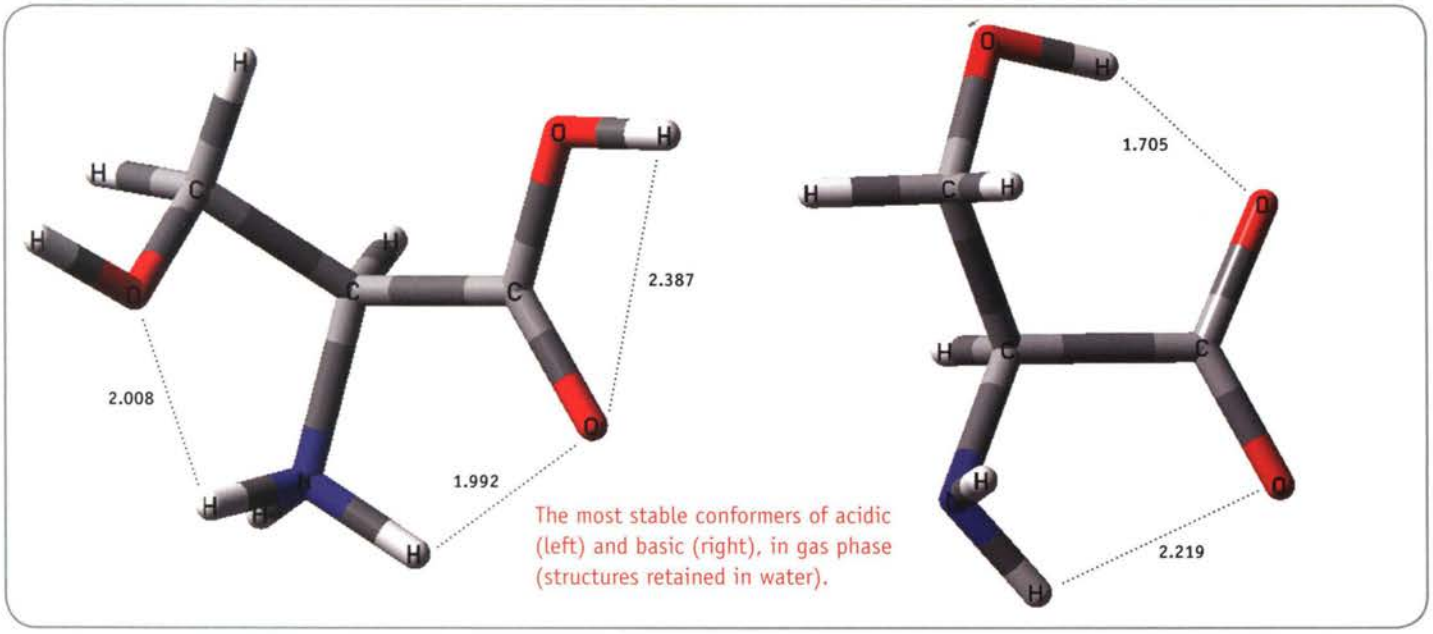
Nurturing researchers

At a time when the science community is stressed to include computational science in a traditional four-year undergraduate college science education, where it is so difficult just to get students through the basics, Bojin's introduction of computational science to her community college students is a bit remarkable. And although it requires a significant time investment on her part, it also meets her need to keep active in research. The computational science she teaches complements what she does for own research.

"Before I came to Queensborough I was doing only research, and it would have been very bad if I had been disconnected completely from research work," she says.

Bojin says overall the students she teaches at QCC "lack a lot of the background, you know, not just for physical chemistry but background in chemistry and math and physics." She tries to give them the main idea about computational chemistry. And those who are interested she works with one on one.

"We don't have enough students whose interest is that sustained to try and do computational modeling with a larger, class-sized group. In fall 2008 I had regular meetings with four students who were interested, but at the end of the semester I ended up with only two of them who wanted to do research. It was kind of disappointing."



The most stable conformers of acidic (left) and basic (right), in gas phase (structures retained in water).

She says she has since found that only two research students at a time is better as she has a heavy teaching load. Once her students have a grasp of the method she then begins having them use the NCSA computers.

"I try to encourage them by giving students projects that are not very advanced in terms of the chemistry that they are using, pretty much molecule projects, things that they can quickly understand. We usually just run things in Gaussian so it is not very fancy," she says. "Then we have little workshops on using the program, and they start on their own making files and running them. They look together at the output. It's a process, so it takes a little while until they get things done," she says.

Bojin says her all her computational students have been motivated by the desire to move on to a good four-year school. They knew that doing research was something that was not only interesting but would give them an edge toward that goal.

"And they love computational chemistry because they have more freedom than they have if they work in the lab," she laughs. "In the lab you have more time constraints whereas if you work on the computer you can be on your own and still do work and finish up your project on Sunday night if you want. You can still do something if you are not here. If they have to travel in the winter break or the summer break, at least they can still do something because they have their laptops."

The Queensborough chemistry department is active in terms of research, notes Bojin. In addition to faculty and student research projects, faculty take students every year to undergraduate research meetings, to regional ACS meetings, and once a year to an ACS national meeting. It is easy for students to have a poster or an oral presentation for that kind of motivator, she says. And it gives them a deadline for their project.

Publication in the works

Bojin says when working with undergraduate students work progresses more slowly, and there is not as much progress as needed to have a publication.

"We've studied five amino acids as thoroughly as we could and now, if we put them together, we could make a case. And hopefully a case that works. The work still needs to be finished up. I hope this summer I can get a manuscript draft completed. We are not the first ones to come up with the idea of studying these things. What we are probably the first ones to do, we are the first ones to systematically study every system, to study every case. Even in previous articles, researchers have studied little bits of each of these systems, but they haven't been put all together. If we look at these five cases we have a little story we can tell about the variations in rigidity of the systems and how they change when you immerse them in solvent. It comes out as a more round story than it was even half a year ago."

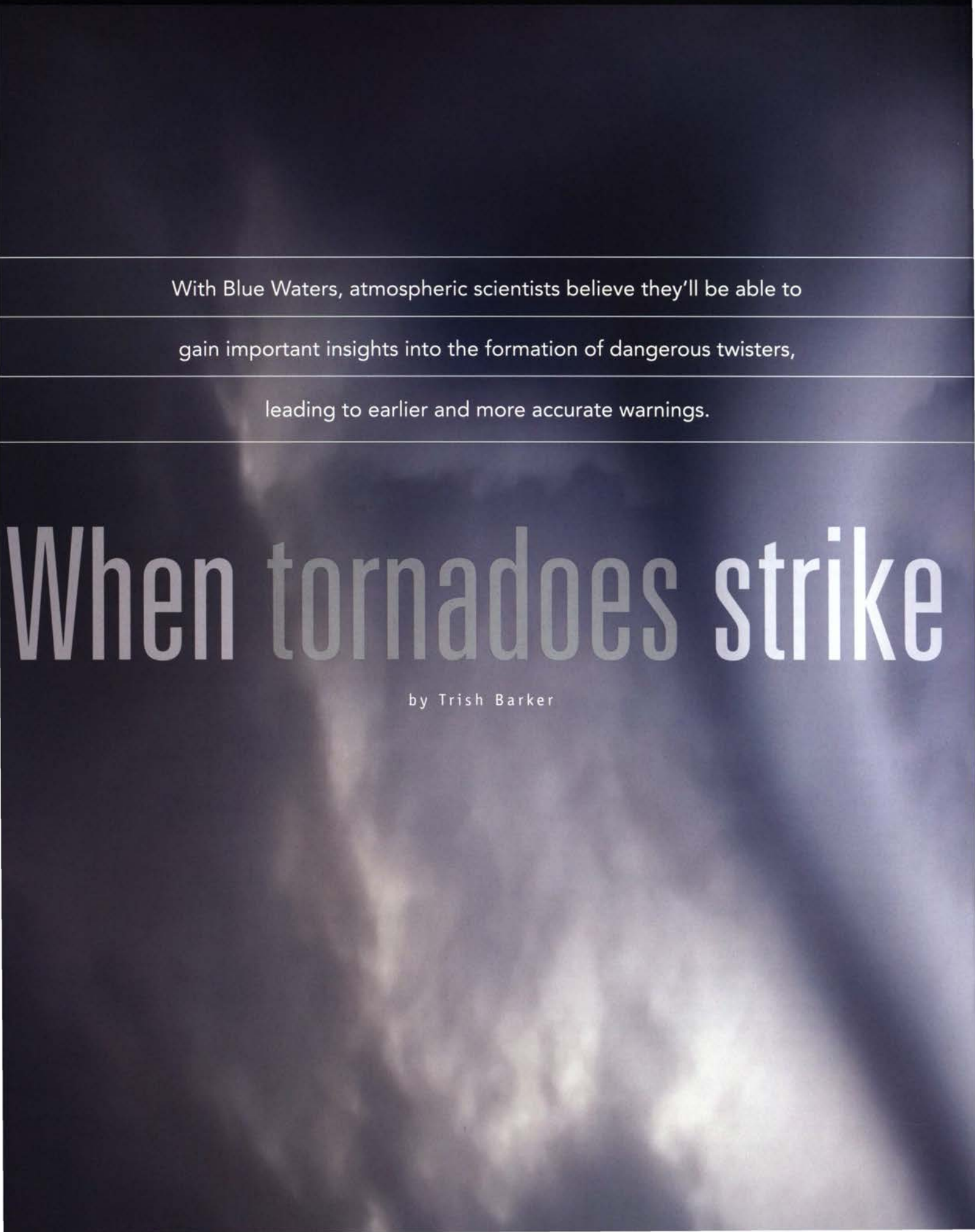
While work with these amino acids is nearing completion, she is working on two more projects with other students. None of it would be possible, she says, without NCSA.

"Given the limited time I have, I don't get a chance to run many calculations so the startup allocation limits are perfect for me. My runs are very little compared to a proper research group where you have things going every day. In my case, my students have something once or twice a week."

But that is more than enough time to inspire them with computational science research. □

Access online

www.ncsa.illinois.edu/News/Stories/inspire



With Blue Waters, atmospheric scientists believe they'll be able to gain important insights into the formation of dangerous twisters, leading to earlier and more accurate warnings.

When tornadoes strike

by Trish Barker

IN MANY WAYS, tornadoes remain a mystery. While their powerful winds can leave broad swaths of devastation—damaged and destroyed homes and businesses, injuries and fatalities—twisters are a challenge for atmospheric scientists to study. It's difficult (not to mention hazardous) to be in the exact right place at precisely the right time to see one in action, and models that capture all of the critical small-scale details within a dynamic storm system are computationally demanding.

That's where the sustained-petaflop Blue Waters supercomputer comes into the picture. The University of Illinois' Bob Wilhelmson is the principal investigator for a Petascale Computing Resource Allocation (PRAC) project. The National Science Foundation's PRAC program enables scientists (in this case Wilhelmson and collaborators at Illinois, the National Center for Atmospheric Research, the National Oceanic and Atmospheric Administration, the University of North Dakota, and Central Michigan University) to work closely with the Blue Waters team, preparing their codes to take full advantage of its several hundred thousand cores.

One target for Wilhelmson's work is tornadogenesis—the process by which a tornado forms. Many tornadoes, including all of the extremely powerful storms that measure EF3, EF4, and EF5 on the Enhanced Fujita Scale, are spawned by supercell thunderstorms. Forecasters can identify conditions in which tornadoes are likely to form, but it's not yet possible to predict exactly when and where tornadoes will form, how strong they will be, or what path they will follow.

"A lot of the key aspects of tornadoes are just not understood," says Brian Jewett, a University of Illinois atmospheric scientist and a member of the PRAC team. "After all these years, how the tornado forms is still a source of considerable controversy."

To address these questions, the PRAC team is working with CM1, a sophisticated model of atmospheric dynamics and thermodynamics, in order to simulate the birth of a tornado from a supercell in the ultra-high resolution needed to adequately represent important small-scale features that influence the evolution of the tornado. For example, without Blue Waters, the team does not believe they could resolve thin curtains of precipitation, tens of meters thick, which potentially transfer large amounts of angular momentum from cloud-base to ground.

Another intriguing small-scale feature is the near-surface inflow layer. "Sometimes you'll see a racing jet of wind a little above the ground that is howling toward the tornado," Jewett says. "This is all going on in the lowest 100 meters above the ground. That's a pretty small zone, but it's a detail that has to be captured. If you don't get these small-scale features, you probably aren't getting the structure of the tornado correct."

Jewett says more detailed simulations using Blue Waters may provide valuable insights into the environmental and particular storm conditions that support tornado development and longevity. "Large, long-lived tornadoes are the ones that cause a lot of the damage and fatalities," he says. "We don't really understand the conditions that allow them to form and to persist. Why do some tornadoes last for an hour or more and others last for seconds?"

Better understanding of what conditions lead to the most dangerous tornadoes can be used to improve the lower-resolution models used in routine forecasting, leading to more precise warnings about the most dangerous twisters.

Getting breakthrough results on Blue Waters won't just be a matter of using existing storm-tornado modeling codes on a more powerful computer. The PRAC team also is working to improve the CM1 model, originally developed by team member George Bryan at the National Center for Atmospheric Research, and the microphysics using the expertise of Matt Gilmore at the University of North Dakota.

"Major modifications to the code are being considered in order to obtain the computational performance necessary to run it at the resolution and with the physics we need to accurately simulate the tornado," Jewett says. "Without them, we will have to either use less sophisticated physics or lower resolution, both of which we do not want to do."

Wilhelmson says it is also necessary for the team to decide what simulations will be run on Blue Waters at higher resolution than ever before and how the vast amount of data generated during these simulations will be analyzed and visualized in a way that improves our understanding and ability to predict severe storms and when tornadoes can be expected to form within them.

"Preparing for Blue Waters is a challenging task that requires the unique expertise of this entire talented team," he says. □

Project at a glance

Team members

George Bryan, National Center for Atmospheric Research
Matthew Gilmore, University of North Dakota
Brian Jewett, University of Illinois at Urbana-Champaign
Joe Klemp, National Center for Atmospheric Research
John Michalakes, National Center for Atmospheric Research
Leigh Orf, Central Michigan University
Glen Romine, National Center for Atmospheric Research
Bill Skamarock, National Center for Atmospheric Research
Lou Wicker, National Oceanic and Atmospheric Administration
Robert Wilhelmson, University of Illinois at Urbana-Champaign (principal investigator)
Paul Woodward, University of Minnesota

Access online

www.ncsa.illinois.edu/News/Stories/PRACtornado

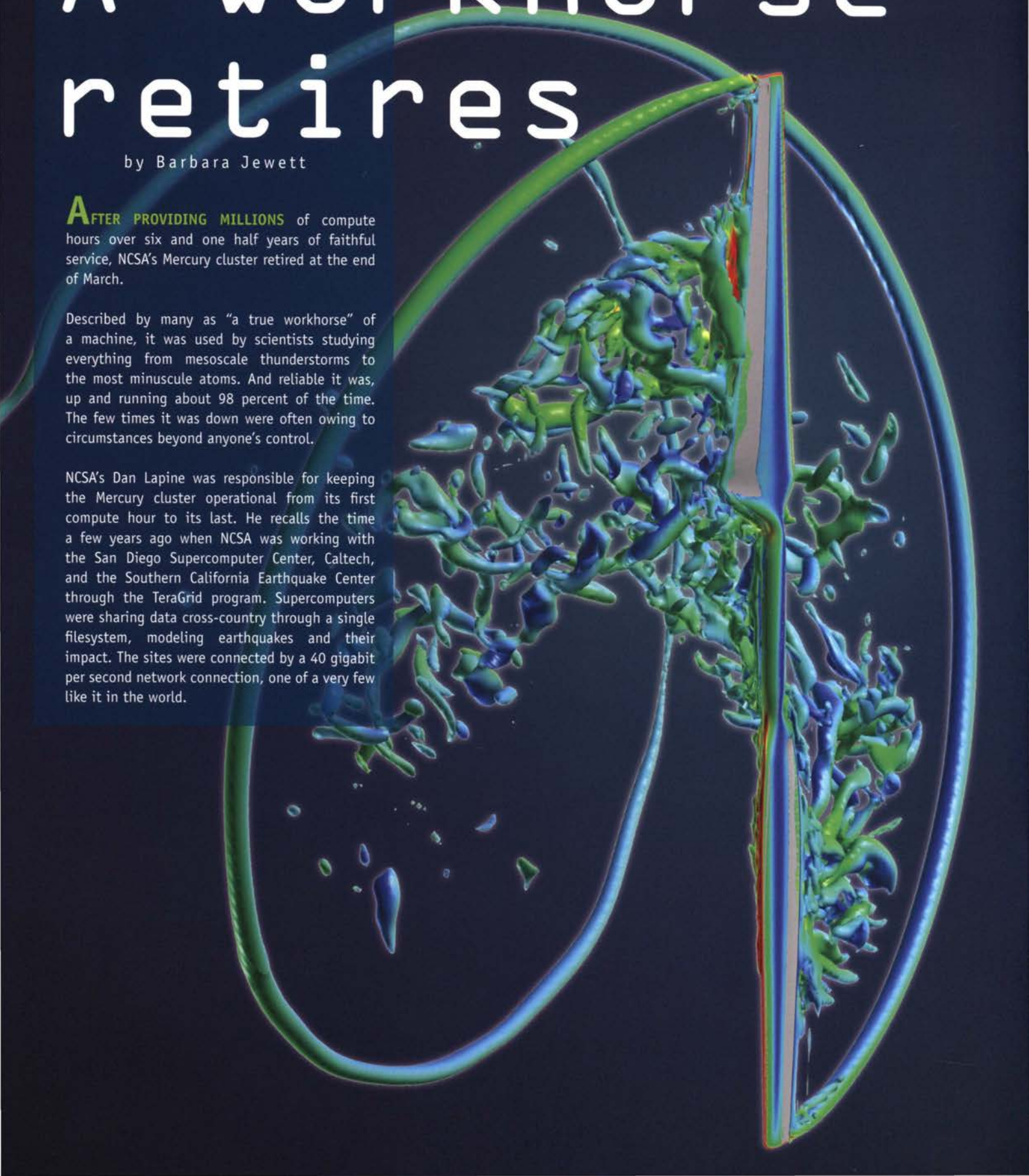
A workhorse retires

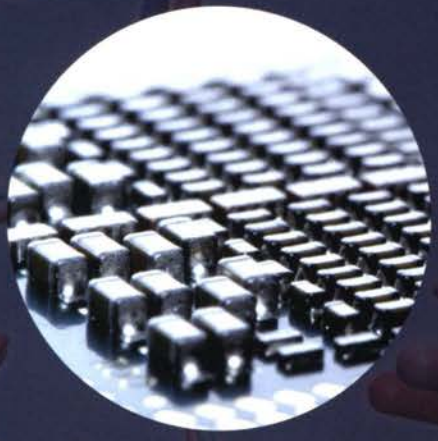
by Barbara Jewett

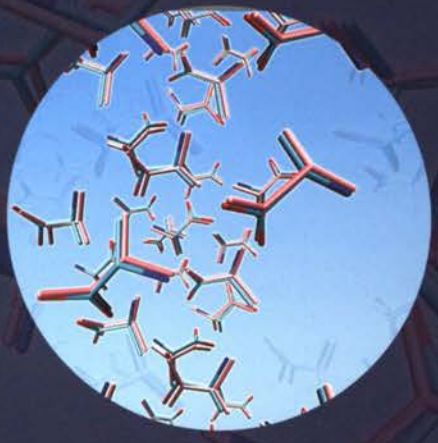
After providing millions of compute hours over six and one half years of faithful service, NCSA's Mercury cluster retired at the end of March.

Described by many as "a true workhorse" of a machine, it was used by scientists studying everything from mesoscale thunderstorms to the most minuscule atoms. And reliable it was, up and running about 98 percent of the time. The few times it was down were often owing to circumstances beyond anyone's control.

NCSA's Dan Lapine was responsible for keeping the Mercury cluster operational from its first compute hour to its last. He recalls the time a few years ago when NCSA was working with the San Diego Supercomputer Center, Caltech, and the Southern California Earthquake Center through the TeraGrid program. Supercomputers were sharing data cross-country through a single filesystem, modeling earthquakes and their impact. The sites were connected by a 40 gigabit per second network connection, one of a very few like it in the world.







The only problem? Networks aren't the only things that travel cross-country. Trains do too.

"Twice—not once—twice, we had train derailments in Colorado where the train tracks parallel the lines that do the data, and the train cut the cable. So we had our file systems crash because of a train derailment in Colorado. A really good example of how our supercomputing and its interruptions don't necessarily depend on us," says Lapine.

Mercury was among the first supercomputers based on the Intel Itanium chip. In fact, one of the test clusters that preceded Mercury had included Itanium chips with serial numbers 1 and 2.

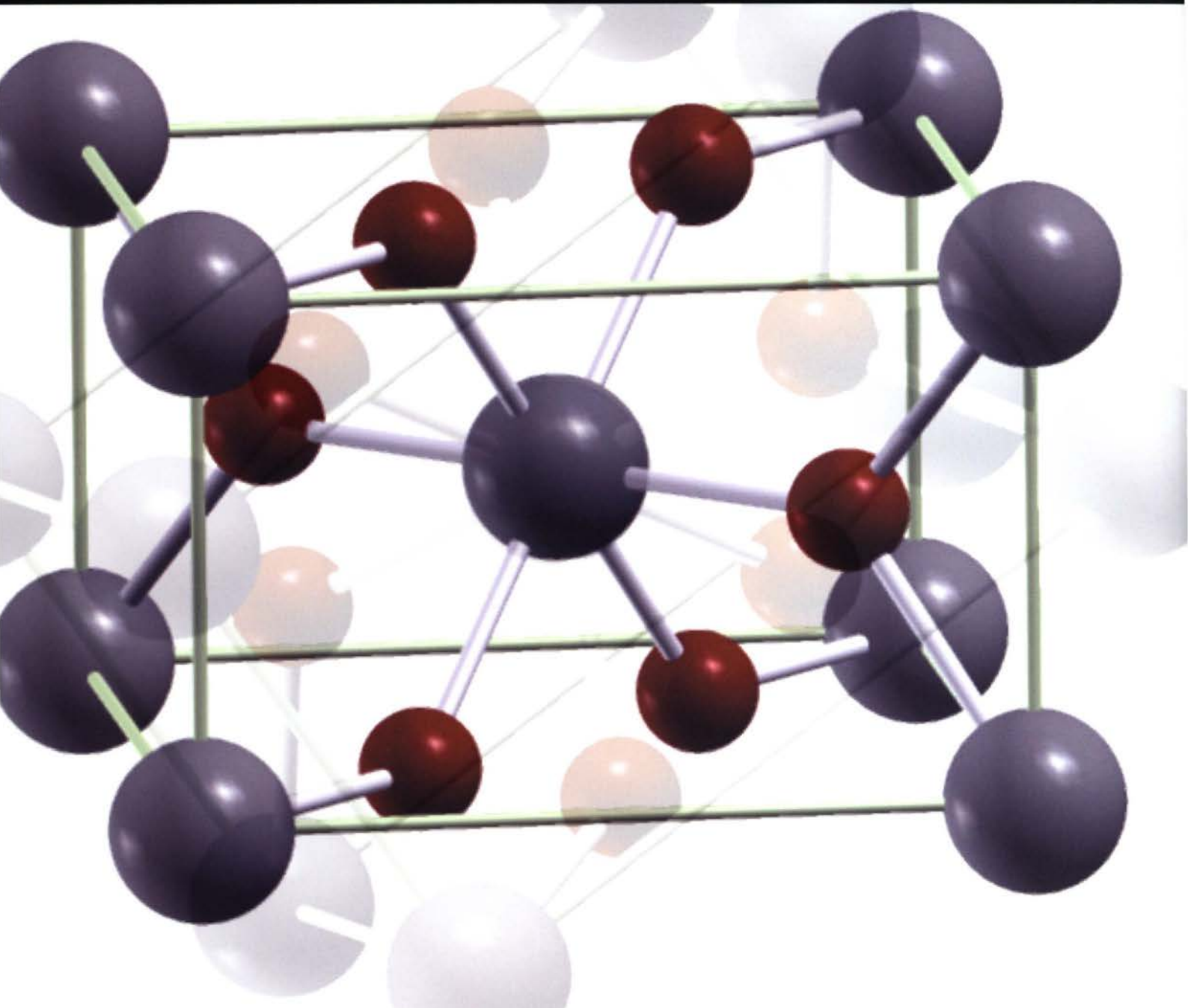
"It was a bit of a risk. No one else was doing it at the time," says Lapine. "It allowed us to have more memory available for each machine and a faster clock speed. But a lot more performance every time the computer would do something. ...When we put the system together, we actually were the 15th fastest computer in the world."

But as technology advanced in the supercomputing world and faster systems arrived on the machine room floor, the goals for Mercury changed. "Over time, the need for Mercury to be the most leading edge became less and less and the need for consistency and stability became more and more," says Lapine. "So being able to keep the machine available 98 percent of the time—I'm quite happy with that."

More than 1,000 researchers had accounts on the system, and more than 2 million jobs were run. Jobs for researchers like Julio Facelli, a professor at the University of Utah who used Mercury to predict crystal structures for organic molecules that are frequently used in pharmaceuticals, fertilizers, and explosives. And for Georgia Tech's Marilyn Smith, who relied on Mercury to model the aerodynamic effects of wind turbines. And for the University of Illinois' Klaus Schulten who depended on Mercury for many of his projects, including researching DNA and gene expression.

These researchers are among the hundreds of Mercury users who conducted transformative research and published papers on the results. For instance, Gautam Ghosh of Northwestern University was one of Mercury's first users. He published nine papers in four years based on work conducted on Mercury. The University of Illinois' Roman Boulatov and his team published eight. In fact, hundreds of papers can be traced back to Mercury's compute power. Not a bad legacy for a revolutionary workhorse of a supercomputer. □

Opposite page and this page, bottom: Marilyn Smith, Georgia Tech. Top: Photo illustration by Blake Harvey, NCSA. Center: Julio Facelli, University of Utah.



Quantum mechanics reveals new details of deep earth

Scientists use NCSA and other TeraGrid resources to reveal that the most common mineral on Earth is relatively uncommon deep within the planet.

by Pam Frost Gorder

SCIENTISTS HAVE USED quantum mechanics to reveal that the most common mineral on Earth is relatively uncommon deep within the planet. Using NCSA's Abe and Cobalt supercomputers, along with LONI's QueenBee, TACC's Ranger and Lonestar, and machines at the Ohio Supercomputer Center, NCAR, NERSC, and the Computational Center for Nanotechnology Innovations, a team of physicists led by Ohio State University has been able to simulate the behavior of silica in a high-temperature, high-pressure form that is particularly difficult to study firsthand in the lab.

The resulting discovery—reported in the online edition of the *Proceedings of the National Academy of Sciences (PNAS)*—could eventually benefit science and industry alike. Silica makes up two-thirds of the Earth's crust, and we use it to form products ranging from glass and ceramics to computer chips and fiber optic cables.

"Silica is all around us," says Ohio State doctoral student Kevin Driver, who led this project for his doctoral thesis. "But we still don't understand everything about it. A better understanding of silica on a quantum-mechanical level would be useful to earth science, and potentially to industry as well."

Silica takes many different forms at different temperatures and pressures —not all of which are easy to study, says Driver. "As you might imagine, experiments performed at pressures near those of Earth's core can be very challenging. By using highly accurate quantum mechanical simulations, we can offer reliable insight that goes beyond the scope of the laboratory. We used the equivalent of 6 million CPU hours or more to model four different states of silica."

Over the past century, seismology and high-pressure laboratory experiments have revealed a great deal about the general structure and composition of the earth. For example, such work has shown that the planet's interior structure exists in three layers called the crust, mantle, and core. The outer two layers—the mantle and the crust—are largely made up of silicates, minerals containing silicon and oxygen.

Still, the detailed structure and composition of the deepest parts of the mantle remain unclear. These details are important for geodynamical modeling, which may one day predict complex geological processes such as earthquakes and volcanic eruptions. Even the role that the simplest silicate—silica—plays in Earth's mantle is not well understood.

"Say you're standing on a beach, looking out over the ocean. The sand under your feet is made of quartz, a form of silica containing one silicon atom surrounded by four oxygen atoms. But in millions of years, as the oceanic plate below becomes subducted and sinks beneath the Earth's crust, the structure of the silica changes dramatically," Driver says.

As pressure increases with depth, the silica molecules crowd closer together, and the silicon atoms start coming into contact with oxygen atoms from neighboring molecules. Several structural transitions occur, with low-pressure forms surrounded by four oxygen atoms and higher-pressure forms surrounded by six. With even more pressure, the structure collapses into a very dense form of the mineral, which scientists call alpha-lead oxide. It's this form of silica that likely resides deep within the earth, in the lower part of the mantle, just above the planet's core, Driver says. When scientists try to interpret seismic signals from that depth, they have no direct way of knowing what form of silica they are dealing with. So they must simulate the behavior of different forms on a computer and then compare the results to the seismic data. The simulations rely on quantum mechanics.

In *PNAS*, Driver, his advisor John Wilkins, and their coauthors describe how they used a quantum mechanical method to design computer algorithms that would simulate the silica structures. When they did, they found that the behavior of the dense, alpha-lead oxide form of silica did not match up with any global seismic signal detected in the lower mantle.

This result indicates that the lower mantle is relatively devoid of silica, except perhaps in localized areas where oceanic plates have subducted, Driver explaines.

Wilkins, Ohio Eminent Scholar and professor of physics at Ohio State, cites Driver's determination and resourcefulness in making this study happen. The physicists used a method called quantum Monte Carlo (QMC), which was developed during atomic bomb research in World War II. To earn his doctorate, Driver worked to show that the method could be applied to studying minerals in the planet's deep interior.

"This work demonstrates both the superb contributions a single graduate student can make, and that the quantum Monte Carlo method can compute nearly every property of a mineral over a wide range of pressure and temperatures," Wilkins says.

Coauthors on the PNAS paper included Ronald Cohen of the Carnegie Institution of Washington; Zhigang Wu of the Colorado School of Mines; Burkhard Militzer of the University of California, Berkeley; and Pablo López Ríos, Michael Towler, and Richard Needs of the University of Cambridge.

Driver notes that the study will "stimulate a broader use of quantum Monte Carlo worldwide to address vital problems." While these algorithms have been around for over half a century, applying them to silica was impossible until recently, he says. The calculations were simply too labor-intensive. Even today, with the advent of more powerful supercomputers and fast algorithms that require less computer memory, the calculations still required using a number of the largest supercomputers in the country.

Driver and his colleagues expect that quantum Monte Carlo will be used more often in materials science in the future, as the next generation of computers goes online. □

This story is adapted from an article that was first published in Research News at The Ohio State University.

Project at a glance

Team members

Kevin Driver
John Wilkens
Ronald Cohen
Zhigang Wu
Burkhard Militzer
Pablo López Ríos
Michael Towler
Richard Needs

Funding

National Science Foundation
Department of Energy

Access online

www.ncsa.illinois.edu/News/Stories/quantum

by Vince Dixon

Ahead of the curve

TYPICALLY, LOCAL CLINICS send reports of illnesses and symptoms to state and national organizations like the Centers for Disease Control. With this method, "the local people who are collecting the data often don't see the results until it is too late to take any action," says Ian Brooks, creator of INDICATOR and director of the Health Sciences Group at NCSA. INDICATOR is an open-source biosurveillance program that analyzes data of local reported illnesses and school absences to detect unusual numbers.

"Although there are existing systems that collect information on infectious diseases, they tend to be top down," Brooks says.

That is why INDICATOR is driven by the local community. It operates from the local-out rather than the standard regional-down approach. This ensures a quicker response time, Brooks says.

It works by collecting data from area schools, hospitals, and clinics. When unusual patterns are detected, the system alerts monitors. This method stresses surveillance to better understand how pandemic diseases can spread and be controlled.

"What we want to do with INDICATOR, because we have computational capability behind it, is be able to link up the surveillance with modeling," Brooks says.

When surveillance and modeling are combined, responses can be planned using actual data as a guide rather than using models alone.

Wendy Edwards is a research programmer at NCSA and member of the Health Science Group. She and Brooks wanted to design INDICATOR so that communities could not only detect disease outbreaks, but also develop possible strategies to address them. Linking surveillance with modeling "would allow practitioners and healthcare providers to decide the best approach at dealing with a disease outbreak," she says.

INDICATOR combines both traditional and syndromic surveillance and uses what its creators call a "cyberenvironment," a system that allows users access to a set of tools that are designed around a specific discipline. The goal is to help researchers design their own workflow and specify the data and algorithms they prefer, Edwards says.

INDICATOR also analyzes data from effectively any kind of source. While other systems require preconceived formats that users input, Brooks and Edwards have written scripts that allow INDICATOR to automatically parse any kind of data format and retrieve the information the team wants. They hope this will encourage more organizations to use the program.

"We want to make it easy for them to work with us," Edwards says.

When local schools, hospitals, and the public health district send data, the analysis algorithm WSARE checks for unusual events. If it detects unusual and meaningful patterns, it sends an alert to the INDICATOR team members and to the public health district. The health district then plans accordingly.

When the H1N1 virus reached Champaign, INDICATOR helped the public health district to better keep track of the number of students absent from schools. With a shortage of vaccinations, the system alerts helped the organization plan for the prioritized distribution of the vaccines. Seeing the number of student absences, the agency revised its distribution strategy to address the apparent increase in outbreak among school children, says Awais Vaid, the epidemiologist for the Champaign-Urbana Public Health District.

"Our goal is that by the time all the data sources start coming in and are analyzed well, it will buy us at least a few days of time before the actual outbreak hits the community," Vaid says. "This system will give us some time to alert community providers to prepare ourselves for something abnormal so that we can plan and respond better."

Once INDICATOR starts to receive heavier amounts of data, the group members hope to run the system on supercomputers.

"When we start getting enough data to justify it, we definitely want to add high-performance computing capabilities to this," Edwards says.

Other plans for IDICATOR include making the system nationally available, Brooks says. And given the close relationship between the spread of human diseases and outbreaks in animal disease, someday INDICATOR will be able to track data from local veterinarians and the University of Illinois College of Veterinarian Medicine, Brooks says. The system's data could also be used for research.

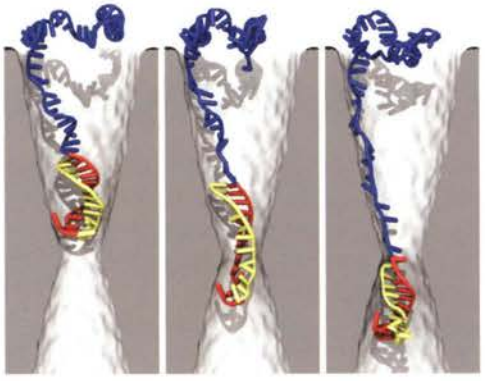
"One of the things you realize quickly when you start working, particularly in the public health area, is that the data you are collecting to make operational decisions is the same data you want to use for research," Brooks says. "It doesn't have to be collected twice." □

Simulations of Nanoscale Biomolecular Systems

Aleksei Aksimentiev, *University of Illinois at Urbana-Champaign*

Nanoscale pores formed in silicon-based materials can be powerful tools for assaying single molecules of DNA, a process with numerous applications in biotechnology and medicine. Professor Aleksei Aksimentiev of the University of Illinois at Urbana-Champaign uses millions of hours of computing time on TeraGrid resources to explore the physics and biology of DNA systems and further his research into utilizing nanopores for DNA sequence detection. His work is part of the race to develop technology and methods to drive the cost of decoding a human genome down to as little as $1,000 by the end of 2013.

Through molecular dynamics (MD) simulations, Aksimentiev and graduate student Jeffrey Comer showed that DNA could pass through such nanoscale pores in different ways depending on the diameter of the pores. Furthermore, experimentalists observed both large increases and large decreases in the current of ions through the pores due to the presence of DNA. The team's simulations elucidated how different positionings of the DNA inside the pore could lead to such large increases and decreases of the current.



Snapshots from a molecular dynamics simulation of a DNA molecule passing through a synthetic nanopore. The double helix of the molecule stretches as it passes through the pore's constriction.

Their findings were based on MD simulations of DNA-nanopore systems for a total simulation time of more than 900 nanoseconds. Nearly half of this simulation time was obtained on the NCSA supercomputers Cobalt and Abe; TACC's Ranger was also used. Furthermore, the NCSA Mass Storage System continues to house large MD trajectories produced by these simulations. This work was published in *Biophysical Journal* in 2009.

This work is supported by grants from the National Institutes of Health, the National Science Foundation, and the Petroleum Research Fund.

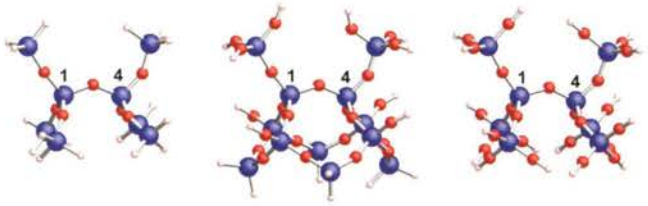
Studying J Couplings

Bradley Chmelka, *University of California, Santa Barbara*

The research team of Bradley Chmelka is motivated by the need to understand at a molecular level the fabrication and functions of new catalysts, adsorbents, porous ceramics, and heterogeneous polymers, technologically important materials linked by their crucial dependencies on local order/disorder, which often govern macroscopic process or device performance. The team is observing many common molecular features among these diverse systems, which provide new insights for materials chemistry and engineering.

Using NCSA's Cobalt, two-bond scalar $^{2}J(^{29}\text{Si-O-}^{29}\text{Si})$ spin-spin couplings between ^{29}Si atoms connected via bridging oxygen atoms in complicated nanoporous zeolite frameworks were calculated with high accuracies by density functional theory (DFT) and compared against experimentally measured values obtained from state-of-the-art two-dimensional $^{29}\text{Si}\{^{29}\text{Si}\}$ nuclear magnetic resonance (NMR) spectroscopy. Such calculations, based on single-crystal or powder X-ray diffraction (XRD) structures of siliceous zeolites Sigma-2 and ZSM-12 and a surfactant-templated layered silicate, were determined to be extremely sensitive to local framework structures. The calculations were compared against experimentally measured couplings obtained with the highest accuracies yet achieved for zeolitic solids using state-of-the-art solid-state 2D double-quantum $^{29}\text{Si}\{^{29}\text{Si}\}$ NMR. The use of $^{2}J(^{29}\text{Si-O-}^{29}\text{Si})$ couplings to probe, evaluate, and refine the local structures of zeolites and layered silicates was demonstrated, including for frameworks that had previously been unsolved.

Such analyses enable new opportunities for the development of protocols that integrate new and sensitive *J* coupling constraints into structure-determination or -refinement protocols. *J* couplings (not restricted to ^{29}Si) are expected to be sensitive probes of local features in a broad range of solids with complicated extents of structural order, including molecularly ordered polymers, organic crystals, biomolecules, and inorganic or hybrid solids. Such local order (or disorder) has important influences on the macroscopic adsorption, reaction, mechanical, photophysical, and/or stability properties of these materials, which can now be better understood and correlated with their structures at a more detailed molecular level.



Different O-centered clusters extracted from the structure of zeolite Sigma-2 (as determined by single-crystal XRD analyses) and centered on the 29Si(1)–O–29Si fragment with different bond terminations. Si, O, and H atoms are displayed in blue, red, and white, respectively.

The team's work, in collaboration with Sylvian Cadars (now at the Centre National de la Recherche Scientifique in Orléans, France) and Darren Brouwer at the National Research Council of Canada, was presented at several conferences and has been published in *Physical Chemistry Chemical Physics* in 2009 and submitted to the *Journal of the American Chemical Society* in 2010.

This work is funded by the U.S. Department of Energy, the National Science Foundation, and Chevron USA.

NCSA a partner in NSF XD TIS award

NCSA, the Texas Advanced Computing Center (TACC), Pittsburgh Supercomputing Center (PSC), and the National Institute for Computational Sciences (NICS) are offering the first dedicated technology insertion service for the "new" TeraGrid. They are partners in the National Science Foundation's (NSF) eXtreme Digital (XD) Technology Insertion Service (TIS) award, an $8.9 million, five-year project commissioned by the Office of Cyberinfrastructure (OCI) to evaluate and recommend new technologies for high-performance computing systems and other resources as part of the NSF TeraGrid and its follow-on initiative, XD.

Since 2005, these centers have contributed to open science and research education by hosting some of the largest HPC resources on the TeraGrid and providing expert technical staff. The XD program will be replacing the TeraGrid program, which provides management of the NSF high-performance computing facilities. The TIS team will be responsible for testing and evaluating the software which will become part of the fabric of the new XD program.

In April 2011, NSF OCI will officially transition from TeraGrid to XD. The TIS award is part of the phased transition process that is taking place now and through April 2011. It plays a critical role in XD's overall mission to accelerate open scientific discovery and to enable researchers to conduct transformational science with next-generation, high-end digital services. John Towns of NCSA is the principal investigator on the TIS grant.

In addition to sustaining continuous improvement in XD's architecture and services, the TIS team will develop and maintain an open, web-accessible database of technology projects for XD sites and users. This database will enable new opportunities for collaboration, research and development, and outreach. The team will also operate the Technology Evaluation Laboratory to ensure that proposed technology changes are thoroughly tested before being recommended for insertion into the production infrastructure.

Up-to-date information, including a TeraGrid to XD transition schedule and answers to frequently asked questions, can be found at: www.teragrid.org/XDTransition.

Snir named Rock Star of HPC

Marc Snir was named an HPC Rock Star by InsideHPC in June. During his time at IBM, Snir contributed to one of the most successful bespoke HPC architectures of the past decade, the IBM Blue Gene. He was also a major participant in the effort to create the most successful parallel programming interface ever: MPI.

With a legacy of success in his portfolio, he is perhaps busier today than ever as co-PI for the petascale Blue Waters system, and co-director of the Intel and Microsoft-funded Universal Parallel Computing Research Center (UPCRC). Trained as a mathematician, Snir is one of the few individuals today shaping both high-end supercomputing and the mass adoption of parallel programming. See the article at http://insidehpc.com/2010/06/10/rock-stars-of-hpc-marc-snir.

More than 1,000 visit new supercomputing facility

More than 1,000 people visited the University of Illinois' new National Petascale Computing Facility July 17, touring the state-of-the-art building and learning more about the supercomputers it will house. Next year, the building will be home to Blue Waters, a supercomputer capable of performing 10 quadrillion calculations every second. Scientists will use Blue Waters to better understand a wide range of phenomena, from the formation of tornadoes to the complex workings of cells in our body. Although the supercomputer is still being built and won't be installed until next year, IBM provided a sample POWER7 node so that visitors could see what the new technology will look like.

Craig, Leetaru receive third web tool patent

NCSA research scientist Alan Craig and colleague Kalev Leetaru have received a third patent for high-powered web tools and methods. Craig is the associate director for human computer interaction for the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) and a researcher at NCSA. Leetaru is a center affiliate.

Their patent, "Method and System for Retrieving Information using an Authentication Web Page," simplifies the process of accessing multiple secure web applications. Instead of managing and re-entering numerous and ever-changing passwords to log onto each application, "you simply browse to a single web page and that takes care of it all," Craig says.

To the password holder, the process is simple. The first time the system is used, the user is given an email or link to a special authentication page. Behind the scenes, this page points the user's web browser to an authentication server and establishes the user's identity. The authentication information is then recorded so each time the user accesses a secured web application, the browser transparently connects to the authentication server, retrieves the latest access privileges for the user on that application, and grants appropriate access. The user never takes further authentication-related actions after accessing the authentication page. Craig and Leetaru have designed a collection of web tools to help everyday web users access advanced web tools without needing advanced know-how.

News & Notes

I-CHASS gets $300,000 grant

The Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) announced that the project Augmented Reality for Understanding Social and Environmental Science has received $300,000 in funding through an EArly-concept Grants for Exploratory Research (EAGER) award from the National Science Foundation's Arctic Social Sciences Program.

Augmented Reality for Understanding Social and Environmental Science will explore and develop the use of augmented reality applications on mobile computing devices to allow students to participate in simulated archeological digs. Augmented reality (AR) is an emerging technology that overlays geographically registered three-dimensional computer graphics on the real world. Recent improvements in cellular telephone technology have enabled the possibility to use a smart phone as an interaction device for AR, thus creating a portable AR experience. By using the cell phone as a "magic lens," participants can view the real world through the camera on the phone and see the augmentations in place, in registration with the real world. The award allows the team to investigate and develop the use of augmented reality in the area of archeology and education.

Alan Craig of I-CHASS and NCSA will work closely with co-PI Robert McGrath and senior scientist David Bock, both research scientists at NCSA. Professor Sophia Perdikaris, from the Department of Anthropology and Archaeology at Brooklyn College, City University of New York, as well as students from the Digital Media program at Parkland College in Champaign, Illinois, will also be collaborators on the project.

Ubiquitous Learning book published

"Ubiquitous Learning," edited by Bill Cope and Mary Kalantzis, includes several chapters authored by NCSA staffers or longtime NCSA collaborators. Jim Myers and Robert McGrath authored a chapter on cyberenvironments. They and staffer Alan Craig co-authored a chapter on immersive online learning environments with Institute for Advanced Computing Applications and Technologies researcher Guy Garnett, and former Faculty Fellow Steve Downie. Edee Norman Wiziecki was on the author team that penned a chapter on Access Grid technology, which also included former Faculty Fellow Sharon Tettegah. Eric Jakobsson authored a chapter on ubiquitous knowledge environments to integrate teaching, learning, and research in the fields of biology and chemistry. Vernon Burton provided a chapter on the history of technology in learning with co-authors James Onderdonk and Simon J. Appleford. Former Faculty Fellow Michael Twidale authored a chapter on ubiquitous computing and ubiquitous learning. Former Faculty Fellow Caroline Haythornthwaite is the author of a chapter called "Participatory Transformations." The book was published by the University of Illinois Press and is available through most book retailers.

Brown elected president

The Great Lakes Consortium for Petascale Computation named Maxine Brown their board president at a meeting in May. Brown is associate director of the Electronic Visualization Lab at the University of Illinois at Chicago. Stan Ahalt, director of the University of North Carolina's Renaissance Computing Institute, was voted president-elect. William Punch, director of Michigan State's High Performance Computing Center, was elected secretary/treasurer. Immediate past president John Ziebarth, senior vice president and chief operating officer of the Krell Insititute, will continue to serve on the executive committee, as will Ross Martin Mehl and Marylyne Munas Mehl of Iowa State University, Steve Gordon of the Ohio Supercomputer Center, Tom Jones of the University of Minnesota, Joe Paris of Northwestern University, and Padma Raghavan of The Pennsylvania State University. More information about GLCPC can be found at: www.greatlakesconsortium.org.

NCSA hosts data centers of future workshop

More than 110 professionals from around the globe involved in building and operating data centers participated in a recent workshop hosted by NCSA. "Building the Data Center of the Future 2nd Biennial Workshop: HPC Data Centers" was held June 23-24 at the I-Hotel and Conference Center in Champaign, Illinois. The first day focused on the future of extreme-scale computing, with talks by Michael Ellsworth of IBM and Shaun Harris of Microsoft, as well as an in-depth look into the design and building of the University of Illinois' National Petascale Computing Facility. Christina Page, director of Energy & Climate for Yahoo!, set the tone for the second day when she presented a variety of factors influencing data center design and presented energy-efficient designs, including some resembling chicken coops. View the complete program at http://www.ncsa.illinois.edu/Conferences/DataCenter.

Industry meets academia

Twenty-one of the Fortune 100 companies participated in NCSA's Private Sector Program meeting in May 2010. Current trends in advanced computational modeling and simulation and how it can benefit industry were discussed, as well as the data management issues companies face. With the completion of the National Petascale Computing Facility and the promise of the future Blue Waters Supercomputer, there was a special emphasis on IBM POWER7 computing for business and industry. The complete program can be viewed at www.ncsa.illinois.edu/Conferences/2010Meeting/agenda.

Open for science

Parting Shot

LESS THAN TWO YEARS after breaking ground, the University of Illinois' National Petascale Computing Facility was completed and ready for occupancy in June. The facility will be home for the Blue Waters sustained-petascale supercomputer when it comes online. The facility is a national asset, providing the infrastructure needed for NCSA's supercomputers and other hardware for decades to come. Other research centers and organizations may also run systems based there.

The National Petascale Computing Facility's data center will have room for the Blue Waters supercomputer, an archive that will be as large as 500 petabytes, other supercomputers, and the necessary hardware that goes with them.

The National Petascale Computing Facility's electrical service is world class. Power usage effectiveness—determined by dividing the amount of power entering a data center by the amount of power being delivered to the computer infrastructure inside it—will be about 1.1 to 1.2. A typical data center rating is often between 1.4 to 2.0.

About 70 percent of the year, cooling tanks—built right into the profile of the building—will chill water by evaporating water in the outside air. That means significantly less electricity will be used for cooling water-cooled supercomputers.

The National Petascale Computing Facility will support both air- and water-cooled supercomputers. Water cooling will make the Blue Waters supercomputer substantially more energy efficient. These pipes are part of the water cooling system.

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

1205 West Clark Street
Urbana, IL 61801

www.ncsa.illinois.edu